ISSN (ONLINE):2045-8711
ISSN (PRINT):2045-869X

International Journal of Innovative Technology and Creative Engineering

May 2024
Vol- 14 No:- 5
@ IJITCE Publication

UK: Managing Editor
International Journal of Innovative Technology and Creative Engineering
1a park lane,
Cranford
London
TW59WA
UK

USA: Editor
International Journal of Innovative Technology and Creative Engineering
Dr. Arumugam
Department of Chemistry
University of Georgia
GA-30602, USA.

India: Editor
International Journal of Innovative Technology & Creative Engineering
36/4 12th Avenue,
1st cross St,
VaigaiColony
Ashok Nagar
Chennai, India 600083

Email: editor@ijitce.co.uk

www.ijitce.co.uk

IJITCE PUBLICATION

International Journal of Innovative Technology & Creative Engineering

Vol.14 No.05

May 2024

IJITCE
INNOVATE TO CREATE

www.ijitce.co.uk

Dear Researcher,

Greetings!

Articles in this issue discusses about study endeavors to recent trends in iris image.

We look forward many more new technologies in the next month.

Thanks,
Editorial Team
IJITCE

Editorial Members

Review Board Members

Contents

DATA SCIENCE IN HEALTH CARE

Dr. P. GOWSALYA
Associate Professor and Head of Biochemistry,
Navarasam Arts and Science college for Women, Arachalur, Erode Tamil Nadu, India
J.BIRUNTHA
Lab Assistant,
Navarasam Arts and Science college for Women, Arachalur,Erode Tamil Nadu, India
S.UDHAYARANI
Lab Assistant,
Navarasam Arts and Science college for Women, Arachalur,Erode Tamil Nadu, India

Abstract- Data science is an interdisciplinary field that extracts knowledge and insights from many structural and unstructured data, using scientific methods, data mining techniques, machine-learning algorithms, and big data. The healthcare industry generates large datasets of useful information on patient demography, treatment plans, results of medical examinations, insurance, etc. The data collected from the Internet of Things (IoT) devices attract the attention of data scientists. Data science provides aid to process, manage, analyze and assimilate the large quantities of fragmented, structured and unstructured data created by healthcare systems. This data requires effective management and analysis to acquire factual results. The process of data cleansing, data mining, data preparation, and data analysis used in healthcare applications is reviewed and discussed in the article. The article provides an insight into the status and prospects of big data analytics in healthcare, highlights the advantages, describes the frameworks and techniques used, briefs about the challenges faced currently and discusses viable solutions. Data science and big data analytics can provide practical insights and aid in the decision-making of strategic decisions concerning the health system. It helps build a comprehensive view of patients, consumers, and clinicians. Data-driven decision-making opens up new possibilities to boost healthcare quality.

Keywords: Big data, Data analytics, Data mining Healthcare, Healthcare informatics.

I. INTRODUCTION

The evolution in the digital era has led to the confluence of healthcare and technology resulting in the emergence of newer data-related applications [1]. Due to the voluminous amounts of clinical data generated from the health care sector like the Electronic Health Records (EHR) of patients, prescriptions, clinical reports, information about the purchase of medicines, medical insurance-related data, investigations, and laboratory reports, there lies an immense opportunity to analyze and study these using recent technologies [2]. The huge volume of data can be pooled together and analyzed effectively using machine-learning algorithms. Analyzing the details and understanding the patterns in the data can help in better decision-making resulting in a better quality of patient care. It can aid to understand the trends to improvise the outcome of medical care, life expectancy, early detection, and identification of disease at an initial stage and required treatment at an affordable cost [3]. Health Information Exchange (HIE) can be implemented which will help in extracting clinical information across various distinct repositories and merge it into a single person's health record allowing all care providers to access it securely. Hence, the organizations associated with healthcare must attempt to procure all the available tools and infrastructure to make use of the big data, which can augment the revenue and profts and can establish better healthcare networks, and stand apart to reap significant benefits [4, 5]. Data mining techniques can create a shift from conventional medical databases to a knowledge-rich, evidence-based healthcare environment in the coming decade.

Big data and its utility in healthcare and medical sciences have become more critical with the dawn of the social media era (platforms such as Facebook and Twitter) and smartphone apps that can monitor personal health parameters using sensors and analyzers [6, 7]. The role of data mining is to improvise the stored user information to provide superior treatment and care. This review article provides an insight into the advantages and

methodologies of big data usage in health care systems. It highlights the voluminous data generated in these systems, their qualities, possible security related problems, data handling, and how this analytics support gaining significant insight into these data set.

Search strategy

A non-systematic review of all data science, big data in healthcare-related English language literature published in the last decade (2010–2020) was conducted in November 2020 using MEDLINE, Scopus, EMBASE, and Google Scholar. Our search strategy involved creating a search string based on a combination of keywords. They were: “Big Data,” “Big Data Analytics,” “Healthcare,” “Artificial Intelligence,” “AI,” “Machine learning,” “ML,” “ANN,” “Convolutional Networks,” “Electronic Health Records,” “EHR,” “EMR,” “Bioinformatics,” and “Data Science.” We included original articles published in English.

Inclusion criteria

Articles on big data analytics, data science, and AI. 2. Full-text original articles on all aspects of application of data science in medical sciences.

Exclusion criteria

1. Commentaries, reviews, and articles with no full-text context and book chapters.
2. Animal, laboratory, or cadaveric studies. The literature review was performed as per the abovementioned strategy. The evaluation of titles and abstracts, screening, and the full article text was conducted for the chosen articles that satisfied the inclusion criteria. Furthermore, the authors manually reviewed the selected article’s references list to screen for any additional work of interest. The authors resolved the disagreements about eligibility for a consensus decision after discussion.

Medical care as a repository for big data

Healthcare is a multilayered system developed specifically for preventing, diagnosing, and treating diseases. The key elements of medical care are health practitioners (physicians and nurses), healthcare facilities (which include clinics, drug delivery centers, and other testing or treatment technologies), and a funding agency that funds the former. Health care practitioners belong to different fields of health such as dentistry, pharmacy, medicine, nursing, psychology, allied health sciences, and many more.

Depending on the severity of the cases, health care is provided at many levels. In all these stages, health practitioners need different forms of information such as the medical history of the patient (data related to medication and prescriptions), clinical data (such as data from laboratory assessments), and other personal or private medical data. The usual practice for a clinic, hospital, or patient to retain these medical documents would be maintaining either written notes or in the form of printed reports [11].

The clinical case records preserve the incidence and outcome of disease in a person’s body as a tale in the family, and the doctor plays an integral role in this tale [12]. With the emergence of electronic systems and their capacity, digitizing medical exams, health records, and investigations is a common procedure today. In 2003, the Institute of Medicine, a division in the National Academies of Sciences and Engineering coined the term “Electronic Health Records” for representing an electronic portal that saves the records of the patients. Electronic health records (EHRs) are automated medical records of patients related to an individual’s physical/mental health or significant reports that are saved in an electronic system and used to record, send, receive, store, retrieve, and connect the medical personnel and patient with medical services [13].

Open-source big data platforms

It is an inefficient idea to work with big data or vast volumes of data into storage considering even the most powerful computers. Hence, the only logical approach to process large quantities of big data available in a complex form is by spreading and processing it on several parallel connected nodes.
Nevertheless, the volume of the data is typically so high that a large number of computing machines are needed in a reasonable period to distribute and finish processing. Working with thousands of nodes involves coping with issues related to paralleling the computation, spreading of data, and manage failures. Table 1 shows the few open sources of big data platforms and their utilities for data scientists.

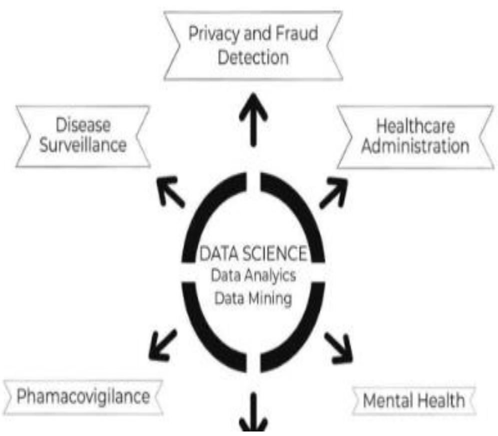
Table 1 Open source big data platforms and their utilities

Big data tools	Utilities
Apache Hadoop	It is designed to scale up to thousands of machines from single servers, each of which offers local storage The framework enables users to easily build and validate distributed structures, distributes data, and operates across machines automatically
Apache Spark	The Hadoop Distributed File system (HDFS) and other data stores are flexible to work with Spark offers integrated Application Program Interfaces (APIs) which enable users to write apps in different languages
Apache Cassandra	Cassandra is highly flexible and can add additional hardware that can handle more data and users on demand Cassandra adapts to all possible data types such as unstructured, structured, and semi-structured supporting features such Atomicity, Consistency, Isolation, and Durability (ACID)
Apache Storm	In several cases, Apache Storm is easy to integrate with any programming language, with real-time analytics, online machine learning, and computation Apache Storm uses parallel calculations which run across a machine cluster
RapidMiner	RapidMiner provides a variety of products for a new process of data mining It provides an integrated data preparation environment, machine learning, text mining, visualization, predictive analysis, application development, prototype validation, and implementation. statistic modeling, deployment
Cloudera	Users can spin clusters, terminate them, and only pay for what they need Cloudera Enterprise can be deployed and run on AWS and Google Cloud Platforms by users

DATA MINING

Data types can be classified based on their nature, source, and data collection methods [14]. Data mining techniques include data grouping, data clustering, data correlation, and mining of sequential patterns, regression, and data storage. There are several sources to obtain healthcare-related data (Fig. 1). The most commonly used type (77%) is the data

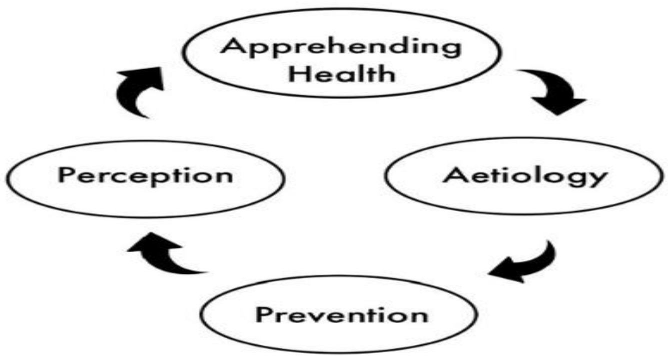
1 Sources of big data in healthcare



generated by humans (HG data) which includes Electronic Medical Records (EMR), Electronic Health Records (EHR), and Electronic Patient Records (EPR). Online data through Web Service (WS) is considered as the second largest form of data (11%) due to the increase in the number of people using social media day by day and current digital development in the medical sector [15]. Recent advances in the Natural Language Processing (NLP)-based methodologies are also making WS simpler to use [16]. The other data forms such as Sensor Data (SD), Big Transactional Data (BTD), and Biometric Data (BM) make around 12% of overall data use, but wearable personal health monitoring devices' prominence and market growth [17] may need SD and BM data.

DISEASE SURVEILLANCE

It involves the perception of the disease, understanding its condition, etiology (the manner of causation of a disease), and prevention (Fig. 3). Information obtained with the help of EHRs, and the Internet has a huge prospect for disease analysis. The various surveillance methods would aid the planning of services, evaluation of treatments, priority setting, and the development of health policy and practice.



Medical signal analytics Telemetry and the devices for the monitoring of physiological parameters generate large amounts of data. The data generated generally are retained for a shorter duration, and thus, extensive research into produced data is neglected. However, advancements in data science in the feld of healthcare attempt to ensure better management of data and provide enhanced patient care [20–23]. The use of continuous waveform in health records containing information generated through the application of statistical disciplines (e.g., statistical, quantitative, contextual, cognitive, predictive, etc.) can drive comprehensive care decision-making. Data acquisition apart from an

ingestion-streaming platform is needed that can control a set of waveforms at various fidelity rates. The integration of this waveform data with the EHR's static data results in an important component for giving analytics engine situational as well as contextual awareness. Enhancing the data collected by analytics will not just make the method more reliable, but will also help in balancing predictive analytics' sensitivity and specificity. The signal processing species must mainly rely on the kind of disease population under observation. Various signal-processing techniques can be used to derive a large number of target properties that are later consumed to provide actionable insight by a pre-trained machine-learning model. Such observations may be analytical, prescriptive, or predictive. Such insights can be furthermore built to activate other techniques such as alarms and physician notifications. Maintaining these continuous waveforms–based data along with specific data obtained from the remaining sources in perfect harmony to find the appropriate patient information to improve diagnosis and treatments of the next generation can be a daunting task [24]. Several technological criteria and specifications at the framework, analytical, and clinical levels need to be planned and implemented for the bedside implementation of these systems into medical setups.

DATA STORAGE AND CLOUD COMPUTING

Data warehousing and cloud storage are primarily used for storing the increasing amount of electronic patient-centric data [25, 26] safely and cost-effectively to enhance medical outcomes. Besides medical purposes, data storage is utilized for purposes of research, training, education, and quality control. Users can also extract files from a repository containing the radiology results by using keywords following the predefined patient privacy policy.

COST AND QUALITY OF HEALTHCARE AND UTILIZATION OF RESOURCES

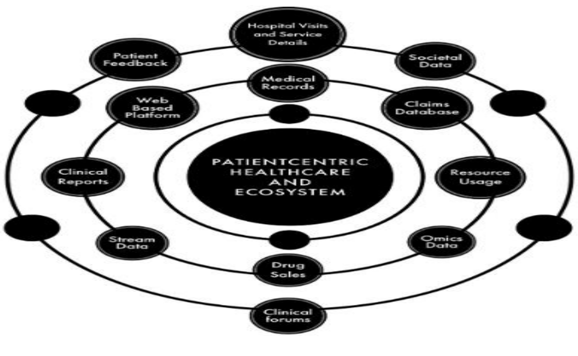
The migration of imaging reports to electronic medical recording systems offers tremendous potential for advancing research and practice on radiology through the continuous updating, incorporation, and exchange of a large volume of data. However, the heterogeneity in how these data can be formatted still poses major challenges. The overall objective of NLP is that the natural human language is translated into structured with a standardized set of value choices that are easily manipulated into subsections or searches for the presence or absence of a finding through software, among other things [27]. Greaves et al. [28] analyzed sentiment (computationally dividing them into categories such as optimistic, pessimistic, and neutral) based on the online response of patients stating their overall experience to predict healthcare quality. They found an agreement above 80% between online platform sentiment analysis and conventional paper-based quality prediction surveys (e.g., cleanliness, positive conduct, recommendation). The newer solution can be a cost-effective alternative to conventional healthcare surveys and studies. The physician's overuse of screening and testing often leads to surplus data and excess costs [29]. The present practice in pathology is restricted by the emphasis on illness. Zhuang et al. [29] compared the disease-based approach in conjunction with database reasoning and used the data mining technique to build a decision support system based on evidence to minimize the unnecessary testing to reduce the total expense of patient care.

Fig. 4 Role of big data in accelerating the treatment process

PATIENT DATA MANAGEMENT

Patient data management involves effective scheduling and the delivery of patient care during the period of a patient's stay in a hospital. The framework of patient-centric healthcare is shown in Fig. 5.

Daggy et al. [30] conducted a study on "no shows" or missing appointments that lead to the clinical capability that has been underused. A logistical regression model is developed using electronic medical records to estimate the probabilities of patients to no-show and show the use of estimates for creating clinical schedules that optimize clinical capacity use while retaining limited waiting times and clinical extra-time. The 400-day clinical call-in process was simulated, and two timetables were developed per day: the conventional method, which assigns one patient per appointment slot, and the proposed method, which schedules patients to balance patient waiting time, additional time, and income according to no-show likelihood. If patient no-show models are mixed with advanced programming approaches, more patients can be seen a day thus enhancing clinical performance. The advantages of implementation of planning software, including certain methodologies, should be considered by clinics as regards no-show costs [30]. A study conducted by Cubillas et al. [31] pointed out that it takes less time for patients who came for administrative purposes than for patients for health reasons. They also developed a statistical design for estimating the number of administrative visits. With a time saving of 21.73% (660,538 min), their model enhanced the scheduling system. Unlike administrative data/target finding patients, a few come very regularly for their medical treatment and cover a significant amount of medical workload. Koskela et al. [32] used both supervised and unsupervised learning strategies to identify and cluster records; the supervised strategy performed well in one cluster with 86% accuracy in distinguishing fare documents from the incorrect ones, whereas the unsupervised technique failed. This approach can be applied to the semi-automate EMR entry system [32].



PRIVACY OF MEDICAL DATA AND FRAUDULENCY DETECTION

The anonymization of patient data, maintaining the privacy of the medical data and fraudulency detection in healthcare, is crucial. This demands efforts from data scientists to protect the big data from hackers. Mohammed et al. [33] introduced a unique anonymisation algorithm that works for both distributed and centralized anonymization and discussed the problems of privacy security. For maintaining data usefulness without the loss of any data privacy, the researchers further proposed a model that performed far better than the traditional K-anonymization model. In addition to this, their algorithm could also deal with voluminous, multi-dimensional datasets.

MENTAL HEALTH

According to National Survey conducted on Drug Use and Health (NSDUH), 52.2% of the total population in the United States (U.S.) was affected by either mental problems or drug addiction/abuse [38]. In addition, approximately 30 million suffer from panic attacks and anxiety disorders [39]. Panagiotakopoulos et al. [40] developed a data analysis–focused treatment technique to help doctors in managing patients with anxiety disorders. The authors used static information that includes personal information such as the age of the individual, sex, body and skin types, and family details and dynamic information like the context of stress, climate, and symptoms to construct static and dynamic information based on user models. For the first three services, relationships between different complex parameters were established, and the remaining one was mainly used to predict stress rates under various scenarios. This model was verified with the help of data collected from twenty-seven volunteers who are selected via the anxiety assessment survey. The applications of data analytics in the disease diagnosis, examination, or treatment of patients with mental wellbeing are very different from using analytics to anticipate cancer or diabetes. In this case, the data context (static, dynamic, or non-observable environment) seems to be more important compared to data volume [39].

PUBLIC HEALTH

Data analytics have also been applied to the detection of disease during outbreaks. Kostkova et al. [43] analyzed online records based on behaviour patterns and media reporting the factors that affect the public as well as professional patterns

of search-related disease outbreaks. They found distinct factors affecting the public health agencies' skilled and layperson search patterns with indications for targeted communications during emergencies and outbreaks. Rathore et al. [44] have suggested an emergency tackling response unit using IoT-based wireless network of wearable devices called body area networks (BANs). The device consists of "intelligent construction," a model that helps in processing and decision making from the data obtained from the sensors. The system was able to process millions of users' wireless BAN data to provide an emergency response in real-time.

PHARMACOVIGILANCE

Pharmacovigilance requires tracking and identification of adverse drug reactions (ADRs) after launch, to guarantee patient safety. ADR events' approximate social cost per year reaches a billion dollars, showing it as a significant aspect of the medical care system [46]. Data mining findings from adverse event reports (AERs) revealed that mild to lethal reactions might be caused in paclitaxel among which docetaxel is linked with the lethal reaction while the remaining 4 drugs were not associated with hypersensitivity [47] while testing ADR's "hypersensitivity" to six anticancer agents [47]. Harpaz et al. [46] disagreed with the theory that adverse events might be caused not just due to a single medication but also due to a mixture of synthetic drugs. It is found that there is a correlation between a minimum of one drug and two AEs or two drugs and one AE in 84% of AERs studies. Harpaz R et al. [47] improved precision in the identifcation of ADRs by jointly considering several data sources. When using EHRs that are available publicly in conjunction with the AER studies of the FDA, they achieved a 31% (on average) increase in detection [45]. The authors identifed dose-dependent ADRs with the help of models built from structured as well as unstructured EHR data [48]. Of the top 5 ADR-related drugs, 4 were observed to be dose-related [49]. The use of text data that is unstructured in EHRs [50]; pharmacovigilance operation was also given priority. ADRs are uncommon in conventional pharmacovigilance, though it is possible to get false signals while finding a connection between a drug and any potential ADRs. These false alarms can be avoided because there is already a list of potential ADRs that can be of great help in potential pharmacovigilance activities [18].

OVERCOMING THE LANGUAGE BARRIER

Having electronic health records shared worldwide can be beneficial in analyzing and comparing disease incidence and treatments in different countries. However, every country would use their language for data recording. This language barrier can be dealt with the help of multilingual language models, which would allow diversified opportunities for Data Science proliferation and to develop a model for personalization of services. These models will be able to understand the semantics — the grammatical structure and rules of the language along with the context — the general understanding of words in different contexts. For example: "I'll meet you at the river bank." "I have to deposit some money in my bank account." The word bank means different things in the two contexts, and a well-trained language model should be able to differentiate between these two. Cross-lingual language model trains on multiple languages simultaneously. Some of the cross lingual language models include: mBERT — the multilingual BERT which was developed by Google Research team. XLM — cross lingual model developed by Facebook AI, which is an improvisation over mBERT. Multift — a QRNN-based model developed by Fast.Ai that addresses challenges faced by low resource language models

CHALLENGES

Millions of data points are accessible for EHR-based phenotyping involving a large number of clinical elements inside the EHRs. Like sequence data, handling and controlling the complete data of millions of individuals would also become a major challenge [51]. The key challenges faced include:

• The data collected was mostly either unorganized or inaccurate, thus posing a problem to gain insights into it.

• The correct balance between preserving patient-centric information and ensuring the quality and accessibility of this data is difficult to decide.

• Data standardization, maintaining privacy, efficient storage and transfers require a lot of manpower to constantly monitor and make sure that the needs are met.

• Integrating genomic data into medical studies is critical due to the absence of standards for producing next-generation sequencing (NGS) data, handling bio

informatics, data deposition, and supporting medical decision-making [52].

• Language barrier when dealing data

FUTURE DIRECTIONS

Healthcare services are constantly on the lookout for better options for improving the quality of treatment. It has embraced technological innovations intending to develop for a better future. Big data is a revolution in the world of health care. The attitude of patients, doctors, and healthcare providers to care delivery has only just begun to transform. The discussed use of big data is just the iceberg edge. With the proliferation of data science and the advent of various data-driven applications, the health sector remains a leading provider of data-driven solutions to a better life and tailored services to its customers. Data scientists can gain meaningful insights into improving the productivity of pharmaceutical and medical services through their broad range of data on the healthcare sector including financial, clinical, R&D, administration and operational details.

CONCLUSION

Larger patient datasets can be obtained from medical care organizations that include data from surveillance, laboratory, genomics, imaging, and electronic healthcare records. This data requires proper management and analysis to derive meaningful information. Long-term visions for self-management, improved patient care, and treatment can be realized by utilizing big data. Data Science can bring in instant predictive analytics that can be used to obtain insights into a variety of disease processes and deliver patient-centric treatment. It will help to improvise the ability of researchers in the fled of science, epidemiological studies, personalized medicine, etc. Predictive accuracy, however, is highly dependent on efficient data integration obtained from different sources to enable it to be generalized. Modern health organizations can revolutionize medical therapy and personalized medicine by integrating biomedical and health data. Data science can effectively handle, evaluate and interpret big data by creating new paths in comprehensive medical care. Funding Open access funding provided by Manipal Academy of Higher Education, ManipAL

DECLARATIONS

Conflict of interest The authors declare no competing interests.

REFERENCES

1. Sengupta PP (2013) Intelligent platforms for disease assessment: novel approaches in functional echocardiography. JACC: Cardiovascular Imaging 6(11):1206–1211. https://doi.org/10. 1016/j.jcmg.2013.09.003
2. Muni Kumar N, Manjula R (2014) Role of big data analytics in rural health care-a step towards svasth bharath. Int J Comp Sci Inform Technol 5(6):7172–7178
3. Ren Y, Werner R, Pazzi N, Boukerche A (2010) Monitoring patients via a secure and mobile healthcare system. IEEE Wirel Commun 17(1):59–65
4. IBM Corporation (2013) Data-driven healthcare organizations use big data analytics for big gains. https://silo.tips/ download/ibm-software-white-paper-data-driven-healthcareorganizations-use-big-data-analy
5. Burghard C (2012) Big data and analytics key to accountable care success. IDC health insights :1–9
6. Bollen J, Mao H, Zeng X (2010) Twitter mood predicts the stock market. J Comp Sci 2(1):1–8. https://doi.org/10.1016/j. jocs.2010.12.007
7. Kuehn BM (2013) NIH recruits centers to lead efort to leverage "big data." JAMA 310(8):787–787
8. Castiglione A, Pizzolante R, De Santis A, Carpentieri B, Castiglione A, Palmieri F (2015) Cloud-based adaptive compression and secure management services for 3D healthcare data. Futur Gener Comput Syst 43:120–134.

9. De Mauro A, Greco M, Grimaldi M (2016) A formal definition of big data based on its essential features. Lib. Review 65(3):122–135.
10. Gubbi J, Buyya R, Marusic S, Palaniswami M (2013) Internet of Things (IoT): a vision, architectural elements, and future directions. Futur Gener Comput Syst 29(7):1645–1660
11. AOCNP, D. (2015) The evolution of the electronic health record. Clin J Oncol Nurs 19(2):153
12. Atasoy H, Greenwood BN, McCullough JS (2019) The digitization of patient care: a review of the efects of electronic health records on health care quality and utilization. Annu Rev Public Health 40:487–500
13. Reisman M (2017) EHRs: the challenge of making electronic data usable and interoperable. Pharmacy and Therapeutics 42(9):572
14. Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare: promise and potential. Health information science and systems 2(1):3
15. Issa NT, Byers SW, Dakshanamurthy S (2014) Big data: the next frontier for innovation in therapeutics and healthcare. Expert Rev Clin Pharmacol 7(3):293–298
16. Baldwin T, Cook P, Lui M, MacKinlay A, Wang L (2013) How noisy social media text, how difrnt social media sources?. In Proceedings of the Sixth International Joint Conference on Natural Language Processing (pp. 356–364)
17. Wang C, Guo X, Wang Y, Chen Y, Liu B (2016) Friend or foe? Your wearable devices reveal your personal pin. In Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security (pp. 189–200)
18. Ventola CL (2018) Big data and pharmacovigilance: data mining for adverse drug events and interactions. Pharmacy and therapeutics 43(6):340
19. Bihan K, Lebrun-Vignes B, Funck-Brentano C, Salem JE (2020) Uses of pharmacovigilance databases: an overview. Therapies 75(6):591–598
20. Seshadri DR, Li RT, Voos JE, Rowbottom JR, Alfes CM, Zorman CA, Drummond CK (2019) Wearable sensors for monitoring the physiological and biochemical profle of the athlete. NPJ digital medicine 2(1):1–16
21. Wang H, Xu Z, Fujita H, Liu S (2016) Towards felicitous decision making: an overview on challenges and trends of big data. Inf Sci 367:747–765
22. Skourletopoulos G et al (2017) Big Data and Cloud Computing: A Survey of the State-of the-Art and Research Challenges. In: Mavromoustakis C, Mastorakis G, Dobre C (eds) Advances in Mobile Cloud Computing and Big Data in the 5G Era. Studies in Big Data, vol Springer, Cham.
23. Yin Z, Lan H, Tan G, Lu M, Vasilakos AV, Liu W (2017) Computing platforms for big biological data analytics: perspectives and challenges. Comput Struct Biotechnol J 15:403–411
24. Elshazly H, Azar AT, El-Korany A, Hassanien AE (2013) Hybrid system for lymphatic diseases diagnosis. In 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 343–347). IEEE.
25. Zhou X, Chen S, Liu B, Zhang R, Wang Y, Li P, Yan X (2010) Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. Artif Intell Med 48(2–3):139–152
26. Yang JJ, Li J, Mulder J, Wang Y, Chen S, Wu H, Pan H (2015) Emerging information technologies for enhanced healthcare. Comput Ind 69:3–11
27. Cai T, Giannopoulos AA, Yu S, Kelil T, Ripley B, Kumamaru KK, Mitsouras D (2016) Natural language processing technologies in radiology research and clinical applications. Radiographics 36(1):176–191
28. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L (2013) Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Int Res 15(11):e239.
29. Zhuang ZY, Wilkin CL, Ceglowski A (2013) A framework for an intelligent decision support system: a case in pathology test ordering. Decis Support Syst 55(2):476–487
30. Daggy J, Lawley M, Willis D, Thayer D, Suelzer C, DeLaurentis PC, Sands L (2010) Using no-show modeling to improve clinic performance. Health Informatics J 16(4):246–259
31. Cubillas JJ, Ramos MI, Feito FR, Ureña T (2014) An improvement in the appointment scheduling in primary health care centers using data mining. J Med Syst 38(8):89
32. Koskela TH, Ryynanen OP, Soini EJ (2010) Risk factors for persistent frequent use of the primary health care services among frequent attenders: a Bayesian approach. Scand J Prim Health Care 28(1):55–61
33. Mohammed N, Fung BC, Hung PC, Lee CK (2010) Centralized and distributed anonymization for high-dimensional healthcare data. ACM Transactions on Knowledge Discovery from Data (TKDD) 4(4):1–33

34. Youssef AE (2014) A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments. Int J Ambient Syst Appl 2(2):1–11
35. Li F, Zou X, Liu P, Chen JY (2011) New threats to health data privacy. BMC Bioinformatics, 12(S12). https://doi.org/10.1186/ 1471-2105-12-s12-s7
36. Joudaki H, Rashidian A, Minaei-Bidgoli B, Mahmoodi M, Geraili B, Nasiri M, Arab M (2015) Using data mining to detect health care fraud and abuse: a review of literature. Global J Health Sci 7(1):194
37. Yang WS, Hwang SY (2006) A process-mining framework for the detection of healthcare fraud and abuse. Expert Syst Appl 31(1):56–68
38. Chong SA, Abdin E, Vaingankar JA, Heng D, Sherbourne C, Yap M, Subramaniam M (2012) A population-based survey of mental disorders in Singapore. Annals of the Academy of MedicineSingapore 41(2):
39. Walker ER, Druss BG (2017) Cumulative burden of comorbid mental disorders, substance use disorders, chronic medical conditions, and poverty on health among adults in the USA. Psychol Health Med 22(6):727–735
40. Panagiotakopoulos TC, Lyras DP, Livaditis M, Sgarbas KN, Anastassopoulos GC, Lymberopoulos DK (2010) A contextual data mining approach toward assisting the treatment of anxiety disorders. IEEE Trans Inf Technol Biomed 14(3):567–581
41. Chen HY, Chuang CH, Yang YJ, Wu TP (2011) Exploring the risk factors of preterm birth using data mining. Expert Syst Appl 38(5):5384–5387
42. Adomavicius G, Tuzhilin A (2011) Context-Aware Recommender Systems. In: Ricci F, Rokach L, Shapira B, Kantor P (eds) Recommender Systems Handbook.
43. Kostkova P, Fowler D, Wiseman S, Weinberg JR (2013) Major infection events over 5 years: how is media coverage infuencing online information needs of health care professionals and the public?. J Med Int Research 15(7):e107. https://doi.org/10.2196/jmir.2146
44. Rathore MM, Ahmad A, Paul A, Wan J, Zhang D (2016) Realtime medical emergency response system: exploiting IoT and big data for public health. J Med Syst 40(12):283
45. Jiang J, Cameron AF, Yang M (2020) Analysis of massive online medical consultation service data to understand physicians' economic return: observational data mining study. JMIR medical informatics 8(2):e16765.
46. Harpaz R, Vilar S, DuMouchel W, Salmasian H, Haerian K, Shah NH, Friedman C (2013) Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. J Am Med Inform Assoc 20(3):413–419
47. Harpaz R, Chase HS, Friedman C (2010) Mining multi-item drug adverse efect associations in spontaneous reporting systems. In BMC bioinformatics (Vol. 11, No. S9, p. S7). BioMed Centra. https://doi.org/10.1186/1471-2105-11-S9-S7
48. Kadoyama K, Kuwahara A, Yamamori M, Brown JB, Sakaeda T, Okuno Y (2011) Hypersensitivity reactions to anticancer agents: data mining of the public version of the FDA adverse event reporting system, AERS. J Exp Clin Cancer Res 30(1):93
49. Akay A, Dragomir A, Erlandsson BE (2014) Network-based modeling and intelligent data mining of social media for improving care. IEEE J Biomed Health Inform 19(1):210–218
50. Eriksson R, Werge T, Jensen LJ, Brunak S (2014) Dose-specifc adverse drug reaction identifcation in electronic patient records: temporal data mining in an inpatient psychiatric population. Drug Saf 37(4):237–247
51. Shvachko K, Kuang H, Radia S, Chansler R (2010) The hadoop distributed fle system. In 2010 IEEE 26th symposium on mass storage systems and technologies (MSST) (pp. 1–10). Ieee. https://doi.org/10.1109/MSST.2010.5496972
52. Gopalani S, Arora R (2015) Comparing apache spark and map reduce with performance analysis using k-means. Int J Comp Appl 113(1)


May 2024
Vol- 14 No:- 5
@ IJITCE Publication